

DESENVOLVIMENTO DA SOLUÇÃO DE BUSINESS INTELLIGENCE EM UMA UNIVERSIDADE ESTADUAL, UTILIZANDO A FERRAMENTA EXCEL PARA MANIPULAÇÃO DOS DADOS

Flávio Augusto Lacerda de Farias*

Rogério Tronco Vassoler**

Resumo

Este trabalho apresenta a utilização das soluções da Microsoft, através da plataforma de Business Intelligence (BI) centralizada no SQL Server. Esse ambiente foi implantado em uma Universidade Estadual, utilizando os dados dos perfis dos candidatos do processo seletivo, com intuito de auxiliar os gestores na tomada de decisão. Foi utilizado o Excel, como ferramenta principal na manipulação e visualização dos dados resultantes, o que tornou a relação custo/benefício viável, visto que, a ferramenta é conhecida e muito utilizada, tornando a curva de aprendizagem menor.

Palavras-chave: Bussiness Intelligence; BI; SQL Server; Excel; Universidade.

Abstract

This work presents the use of Microsoft solutions by business intelligence (BI) platform centered on SQL Server. This environment was implemented at the State University, using data from profiles of candidates in the selection process, aiming to assist managers in decision making. Excel was used as the principal tool in the manipulation and visualization of the resulting data, making the cost / benefit achievable, since the tool, known and widely used, making shorter learning curve.

Keywords: Bussiness Intelligence; BI; SQL Server; Excel; University.

1 INTRODUÇÃO

O processo de tomada de decisão vem evoluindo dentro das organizações, sejam elas públicas ou privadas e a medida que novos desafios surgem oriundos das movimentações econômicas ou políticas, os processos gerenciais são revisados buscando acompanhar as mudanças. Atualmente para uma organização continuar competitiva não basta apenas o investir seu capital em imobilizados ou na mão de obra, faz-se necessário investir principalmente no capital intelectual, conforme Stewart (1998), o capital intelectual é a soma do conhecimento de todos os colaboradores em uma empresa, constituindo a matéria intelectual-conhecimento usada para gerar riqueza. No entanto para que esse capital intelectual seja utilizado em sua plenitude é necessário obtê-los de maneira correta e no momento ideal, pois de nada adianta uma montanha de dados se não forem úteis no momento oportuno.

* Graduando em Sistemas de Informação – Universidade do Salvador – UNIFACS, voluntário na Iniciação Científica no Grupo de Pesquisas em Tecnologia da Informação – GPTI – flavio.augusto.farias@hotmail.com

** Professor Msc. da Universidade do Salvador – UNIFACS, Coordenador do Grupo de Pesquisas em Tecnologia da Informação – GPTI – rogerio.vassoler@pro.unifacs.br

Para cumprir com a crescente demanda, as organizações precisam de ferramentas que as auxiliem de maneira efetiva e imediata, desta forma é necessário que essas ferramentas sejam flexíveis e que possam lidar adequadamente com as constantes mudanças, além disso, devem acima de tudo ser intuitivas para que os tomadores de decisão possam focar nas questões pertinentes a suas posições corporativas, tais sistemas são conhecidos como Sistemas de Suporte a Decisão.

Os usuários dos Sistemas de Suporte a Decisão (SSD), costumam atuar em posições estratégicas dentro da organização, de maneira que toda complexidade técnica deve ser mascarada e encapsulada em soluções práticas e diretas, visto que a informação costuma estar distribuída em diversos sistemas transacionais, planilhas eletrônicas ou arquivos de texto ou até mesmo com o próprio colaborador em meios não digitais de acordo com a IBM (2015) é necessário “construir uma cultura para difundir a análise em todas as áreas da empresa. Permitir que todos os funcionários tomem decisões baseadas em dados, ao invés de confiar no instinto e em experiências prévias”.

Essa cultura pode ser representada sistematicamente pelo Business Intelligence ou BI que é definido pelo Gartner Group (2015) como “um termo abrangente que inclui as aplicações, infraestrutura e ferramentas além das melhores práticas que permitam o acesso e a análise de informações para melhorar e otimizar decisões e desempenhos.”. De maneira que ao separar o BI do ambiente transacional é possível a criação de um repositório central de todos os dados, fazendo com que toda informação necessária possa ser solicitada e acessada diretamente pelo usuário final através de relatórios, gráficos ou alertas. Com isso os usuários podem ter acesso às informações históricas da empresa, permitindo diversas análises detalhadas e tornando o processo de tomada de decisão mais rápido.

2 MINERAÇÃO DE DADOS

De acordo com Kimball (1998) um Data Warehouse armazena uma cópia dos dados transacionais, organizado especificamente para geração de consultas e análises. Para Silberschatz, Korth e Sudarshan (2010) um Data Warehouse é um repositório (ou arquivamento) de informações colhidas de várias origens, armazenadas sob um esquema unificado em um único local por muito tempo.

Para Kimball (1998), um DW deve atingir alguns objetivos:

a) Informação organizada e acessível

Na maioria das empresas, a informação está descentralizada em sistemas transacionais ou em arquivos espalhados por toda organização, então é necessário que todas as informações sejam integradas e com fácil localização.

b) Integridade na informação apresentada

A consistência da informação significa que se duas pessoas acessarem o resultado de um determinado indicador devem obter a mesma resposta, para isso toda informação deve ser consistente.

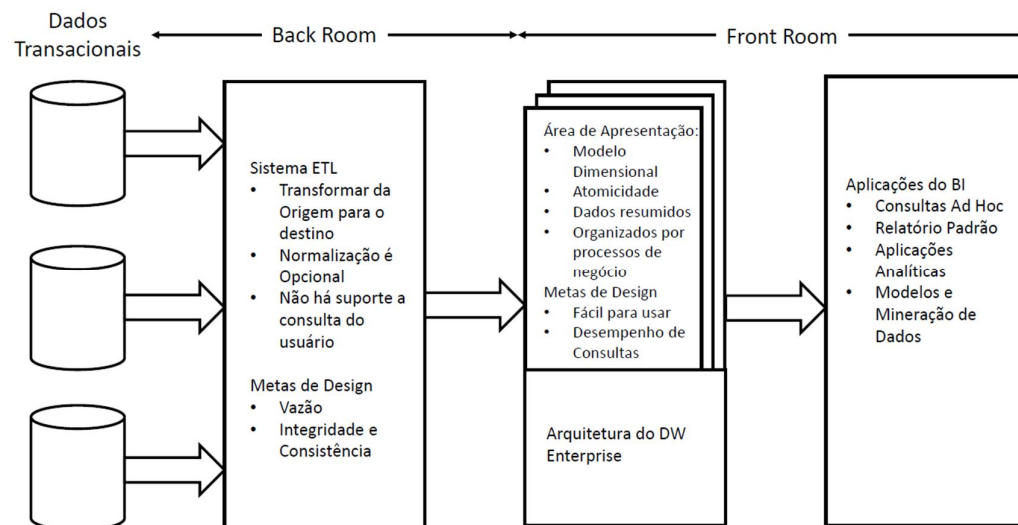
c) Disponibilidade da informação

Para Kimball (1998) 60% do projeto de um data warehouse está relacionado com os componentes “*back room*” tais como hardware, software e dados, os outros 40% representam o “*front room*”, ou seja as ferramentas responsáveis pelas consultas, análises e apresentação dos dados.

3 ARQUITETURA DE UM DATA WAREHOUSE

Na figura 1, temos o exemplo proposto por Kimball (2008) para descrever a arquitetura do ambiente de Data Warehouse.

Figura 1 - Arquitetura DW (Kimball, 2008)



A definição do DW também conhecida por "Armazém de Dados" faz sentido quando analisamos o modelo proposto por Kimball (2008) para descrever a arquitetura de um DW, a origem dos dados são os sistemas transacionais que ficam na camada externa a esse armazém,

é onde toda e qualquer informação é armazenada e manipulada, no entanto, nesse estágio existe pouco ou nenhum controle sobre o conteúdo e o formato dos dados. A principal prioridade desses sistemas é o processamento e a disponibilidade, as consultas ao sistema DW não são implementadas.

Comumente os sistemas transacionais mantêm poucos dados históricos e quando o fazem não armazenam períodos muito longos, nesse ponto um Data Warehouse torna-se peça chave, pois uma das principais vantagens dele é o armazenamento histórico da informação diminuindo assim os acessos aos dados históricos no ambiente transacional que possuem propósito específico. Além dos sistemas transacionais, a informação pode estar diretamente com o usuário em arquivos não estruturados tais como planilhas de texto ou documentos de textos.

Como os dados estão descentralizados e não existe garantia da integridade e consistência da informação é necessário realizar o processo de extração dos dados das diversas fontes, pois somente assim vai ser possível realizar os possíveis ajustes e correções buscando a integridade e consistência dos dados que vão ser carregados para o DW, não é propósito corrigir os sistemas transacionais, mas sim organizar as informações de maneira que os gestores possam ter um ambiente organizado e estruturado para auxiliar na tomada de decisão, esse passo é conhecido como ETL (*Extract, Transformation e Load*) ou seja, extração, transformação e carga dos dados.

Após o processo de ETL os dados são armazenados em estruturas dimensionais modeladas especificamente para esse propósito de maneira que possam ser consultadas pelos usuários. Nesse momento três estruturas se destacam (MACHADO, 2004):

Data Warehouse Empresarial (DWE) – Suporta toda ou a maior parte dos requerimentos ou necessidades além disso é um modelo de DW integrado com grande grau de acesso e utilização das informações para todos os departamentos de uma empresa. Nesse modelo torna-se um repositório comum de dados de suporte a decisão.

Data Mart Independente (DMI) – Basicamente é um DW voltado para um segmento específico de um negócio, sendo implementado de maneira isolada e sem nenhuma relação com as demais áreas da empresa, é possível a existência de outros DMI que da mesma forma são isolados e sem nenhuma relação entre eles.

Data Mart Integrados (DMIT) – São *data marts* oriundos normalmente de um DWE, sendo um modelo híbrido apesar de ser voltado para um segmento específico dentro da organização, ele faz parte de um único DW ou normalmente está integrado a outros DM.

A decisão sobre a arquitetura pode causar diversos impactos no sucesso do projeto, inúmeros fatores influenciam a escolha do modelo a ser implementado entre eles estão o tempo para a execução do projeto, a necessidade de implantação do DW, o retorno do investimento, os recursos necessários à implementação da arquitetura e acima de tudo a satisfação do usuário executivo.

De acordo com Machado (2004), existem vários tipos de implementação de arquiteturas, no entanto três modelos tem grande destaque nos processos de implantação, sendo elas a *top down*, *bottom up* e uma combinação das duas.

A implementação *top down* é considerada como o modelo inicial de Data Warehouse, nesse modelo, o DW é construído integralmente e a partir dele os dados são transferidos para os DM.

4 IMPLEMENTAÇÃO *TOP DOWN*

4.1 Vantagens

Herança de arquitetura: Os DM's compartilham a arquitetura e os dados do DW, o que permite uma fácil manutenção.

Visão de empreendimento: o DW concentra todos os negócios da empresa, sendo possível extrair dele níveis menores de informações.

Controle e Centralização das regras: Garantia da existência de um único conjunto de aplicações para extração, limpeza e integração dos dados.

4.2 Desvantagens

Implementação muito longa: Apesar do DW ser desenvolvido de modo interativo por áreas de assuntos, como, vendas, finanças e RH. Ainda assim demora muito para que a primeira área de assunto entre em produção.

Alta taxa de riscos: Como a apresentação de resultado demora, não existe garantia para o investimento nesse tipo de ambiente, pois o resultado pode não ser satisfatório e o tempo para a recuperação do investimento é muito longo.

Expectativas relacionadas ao ambiente: Como o tempo de apresentação dos resultados são longos, a expectativa em torno do projeto cresce exponencialmente, de maneira que muitas vezes, o resultado apesar de satisfatório parece além do esperado.

5 IMPLEMENTAÇÃO *BOTTOM UP*

A implementação *bottom up* tem como propósito construir um DW de maneira incremental a partir dos diversos DM Independentes, dessa maneira, os resultados são apresentados muito mais rapidamente ao usuário.

Machado (2004) aponta as seguintes vantagens e desvantagens dessa implementação:

5.1 Vantagens

Implementação mais rápida: Como a construção do DM é direcionada, o resultado é apresentado muito mais rapidamente e consequentemente as áreas interessadas obtém resultados em menos tempo.

Retorno rápido: Como o resultado é apresentado em menos tempo, as áreas tomadoras de decisão notam o valor da ferramenta rapidamente, isso permite um investimento maior com mais confiança.

Enfoque nos principais negócios: Como os DM são implementados separadamente, é possível começar pelas áreas mais importantes, mesmo que não sejam as primeiras no modelo de negócio, dessa maneira o gasto é direcionado para as áreas mais imediatas.

5.2 Desvantagens

Dificuldade da visão macro do negócio: Como os DM Independentes são criados normalmente por área de negócio, é importante não perder de vista toda a visão empresarial.

Múltiplas equipes: Como DM são desenvolvidos em paralelo é necessário coordenar diversas equipes e administrar o esforço principalmente nas áreas de regras e semânticas empresariais.

A implementação mista, é planejada e implementada utilizando o modelo *Top Down*, no entanto o desenvolvimento é baseado em *Bottom Up*, de maneira que os DM são entregues um de cada vez, mas cada um deles é visto de maneira evolutiva.

6 MODELAGEM DIMENSIONAL

Kimball (1998) afirma que a modelagem dimensional é um nome novo para uma técnica já difundida na criação de bancos de dados simples, permitindo que um banco de

dados seja visualizado como um cubo contendo três ou mais dimensões possibilitando que os usuários possam “fatiar” esse cubo em relação a qualquer uma dessas dimensões.

Do ponto de vista do um gestor ele descreveria a atividade da empresa da seguinte maneira: “Ao longo do tempo analisamos o total de vendas dos produtos nos diversos mercados”

De acordo com a figura a seguir, temos três dimensões que são (Produto, Mercado e Tempo):

Figura 2 - Modelo dimensional de um negócio: cada ponto interno ao cubo contém as medições para aquela combinação específica de Produto, Mercado e Tempo. (Kimball, 1998)



Cada ponto no cubo está na intersecção das coordenadas e representa a relação entre: Produto – Mercado e Tempo, por exemplo: “10 unidades do produto X foram vendidos ao longo da primeira semana de Janeiro.” Dessa maneira é possível realizar a leitura de maneira simples e fácil e entregar exatamente o que o gestor quer, sem perder tempo com diversas informações, nem sempre o detalhado é o melhor e essa representação simples torna o processo de modelagem do DW eficiente.

O modelo multidimensional é baseado em três tipos de estruturas (Reis; Teixeira; Araújo, 2009):

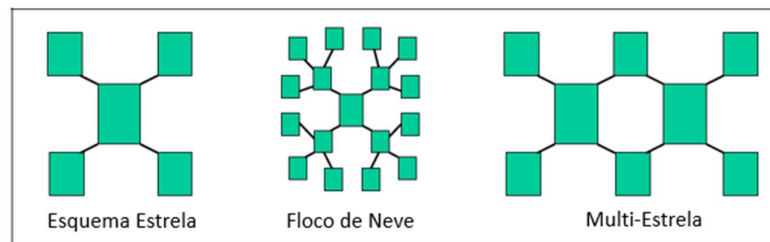
- a) **Fatos** – a tabela de fatos é a tabela central do modelo e contém os valores (numéricos) que se deseja analisar, geralmente, contendo um grande volume de dados. A tabela fato possui chaves externas, que se relacionam com suas tabelas de dimensões, e campos numéricos que são os valores (medidas) que serão analisados;
- b) **Dimensões** – as tabelas de dimensões representam um aspecto do negócio que está sendo analisado. Sua chave primária serve para manter a integridade referencial na tabela fato à qual está relacionada. Uma dimensão oferece ao

usuário um grande número de combinações e intersecções para analisar os dados, possibilitando diversas formas de visualizar os dados;

- c) **Medidas** – são atributos numéricos armazenados na tabela de fatos, que representam o desempenho de um indicador em relação às dimensões que participam desse fato, é o que se deseja de fato conhecer.

Ballard (2006), afirma que existem três modelos básicos multidimensionais conforme figura 3:

Figura 3 - Tipos de modelos de dimensão



- a) **Esquema Estrela** – As dimensões são formadas por apenas uma tabela não normalizada, de maneira que existe uma quantidade menor de tabelas *no* DW, mas em contrapartida pode existir redundância dos dados, aumentando o tamanho das mesmas.
- b) **Esquema Floco de Neve** – Nesse esquema, as tabelas são normalizadas evitando a redundância dos dados. Diferente do esquema estrela nesse esquema os dados estão distribuídos em múltiplas tabelas.
- c) **Esquema Multi-estrela** – Nesse esquema, as tabelas dimensão não estão normalizadas, porém existem múltiplas tabelas de fato, unidas através das dimensões.

7 ANÁLISE E DESENVOLVIMENTO

Este trabalho tem como objetivo desse trabalho foi através das soluções da Microsoft de BI para desenvolver um Data Mart com o perfil dos candidatos do processo seletivo de uma Universidade Estadual, utilizando o Excel como ferramenta principal na manipulação e visualização dos dados resultantes, apoiando e auxiliando a tomada de decisões, no entanto foi realizado um recorte temporal referente aos últimos quatro anos (2010-2014) já que o período anterior a esse o processo de inscrição não era gerido pela IES (Instituição de Ensino Superior), além disso o trabalho também serviu de base para a implantação da solução de BI

dentro da Universidade, englobando diversos outros DM em todos os níveis organizacionais tais como cursos, docentes e discentes.

O desenvolvimento do projeto foi realizado em diversas etapas:

- a) Planejamento
- b) Levantamento das necessidades
- c) Modelagem
- d) Projeto físico do banco de dados
- e) ETL
- f) Desenvolvimento de Aplicações
- g) Homologação e Implantação

A seguir cada um desses passos é detalhado.

8 PLANEJAMENTO

Na etapa de planejamento, foi definida a arquitetura tecnológica do projeto, optou-se pela utilização do banco de dados SQL Server, o *Integration Services* para a realização dos processos de ETL, o *Analysis Services* como ferramenta de modelagem multidimensional e criação dos cubos, além do uso da planilha eletrônica Excel para a criação das visualizações e manipulação dos cubos que vão ser desenvolvidos no decorrer de todo o projeto. A escolha da plataforma e do ambiente da Microsoft foi tomada porque a Universidade tem as suas principais bases armazenadas e administradas no ambiente do SQL Server além de possuir acesso as demais ferramentas utilizadas para demonstrar os recursos voltados para a manipulação de dados multidimensionais utilizando o Excel.

Foi definido nessa etapa a arquitetura e o tipo de implementação que devem ser usados para o desenvolvimento do DW, além da arquitetura do tipo Data Mart Integrado, apenas as informações dos candidatos estão sendo analisadas, utilizando a implementação mista (*top down* e *bottom up*) pois o projeto final para a IES é a solução completa de BI.

A modelagem utilizada para o Data Mart dos Candidatos é o esquema estrela, a escolha desse modelo foi tomada por conta da performance, além disso o objetivo foi diminuir a quantidade de tabelas através da utilização de dados duplicados ou não normalizados.

9 LEVANTAMENTO DAS NECESSIDADES

Nessa etapa foram realizadas reuniões com uma equipe multidepartamental com o intuito apresentar a solução de BI pois a maioria dos participantes que não são da área de tecnologia não estavam ambientados e familiarizados com a ferramenta, além disso foi possível identificar também as principais necessidades para o projeto do BI, pois foi apresentado para os participantes uma matriz de necessidade(figura 5), onde os participantes disponibilizariam as perguntas que gostariam que fossem respondidas, dessa maneira seria possível identificar todas as possíveis dimensões envolvidas na busca da resposta, a utilização da matriz de necessidade foi bem explorada pois foi possível esclarecer diversas dúvidas sobre o projeto.

Figura 5 – Matriz de Necessidades

Matriz de necessidades / UNEB																				
Fato		Curso			Docentes			Discentes			Técnico			Candidato						
Dimensão	Total_Cursos_Departamento_Campus	Total_Cursos_Nivel_Grau-Acadêmico	Total_Cursos_Nivel_Turno_Departame	Total_Cursos_Programas-Especiais_De	Total_Docentes_Departamento_Camp	Total_Docentes_Curso_Departamento	Total_Docentes_Titulação_Departame	Total_Docentes_Titulação_Curso_Depa	Total_Discentes_Cota_TipoEM_Escola	Total_Discentes_Status_Matricula_Cur	Total_Discentes_Genero_curso_Depart	Total_Discentes_Genero_Cota_Curso	Total_Técnicos_Campus	Total_Técnicos_Departamento_Campu	Total_Técnicos_Genero_Departamento	Total_Técnicos_Formação_Departame	Total_Candidatos_Cota_TipoEM_Escol	Total_Candidatos_Gênero_Idade_Depa	Total_Candidatos_Cidade(endereço)-d	Total de Candidatos Cotistas
Campus																				
descrição																				
Departamentos																				
descrição																				
Cursos																				
descrição																				
Docentes																				
descrição																				
Técnicos																				
descrição																				
Discentes																				
descrição																				
Candidatos																				
descrição																				
Tempo																				
descrição																				

O processo de carga das tabelas vai ser baseado nos DM envolvidos, inicialmente como os dados são apenas dos candidatos do vestibular, a inserção dos dados é anual, quando os dados acadêmicos dos discentes forem importados, o processo será baseado no calendário acadêmico, tais como matrícula, avaliações e etc.

Outro requisito importante determina que a solução deve estar operando no modelo 24x7 (Vinte e quatro horas por dia e sete dias por semana), com um acordo de nível de serviço

(*Service Level Agreement* - SLA) de 99,5% de disponibilidade. Todo o hardware e software do servidor será monitorado, pois caso ocorra alguma falha ou instabilidade no sistema o responsável pelo suporte será acionado. Após a finalização do levantamento das necessidades, foi iniciada a modelagem dimensional do DW, conforme apresentada na próxima seção.

10 MODELAGEM

O processo de modelagem dimensional é o mais crítico de toda implementação do BI, as dimensões e tabelas fatos são definidas nesse momento, com o preenchimento da matriz de necessidade, é possível já ter uma noção clara das dimensões usadas em cada ponto do projeto. No DM dos Candidatos, optou-se pelo desmembramento da dimensão para que otimizasse as consultas futuras, três dimensões foram definidas para o candidato, cada uma delas contemplava uma faceta do candidato.

Dimensão Candidato: Informações pessoais referentes ao candidato, tais como: sexo, etnia indígena, tipo do ensino médio e fundamental, se possui algum benefício social e etc.

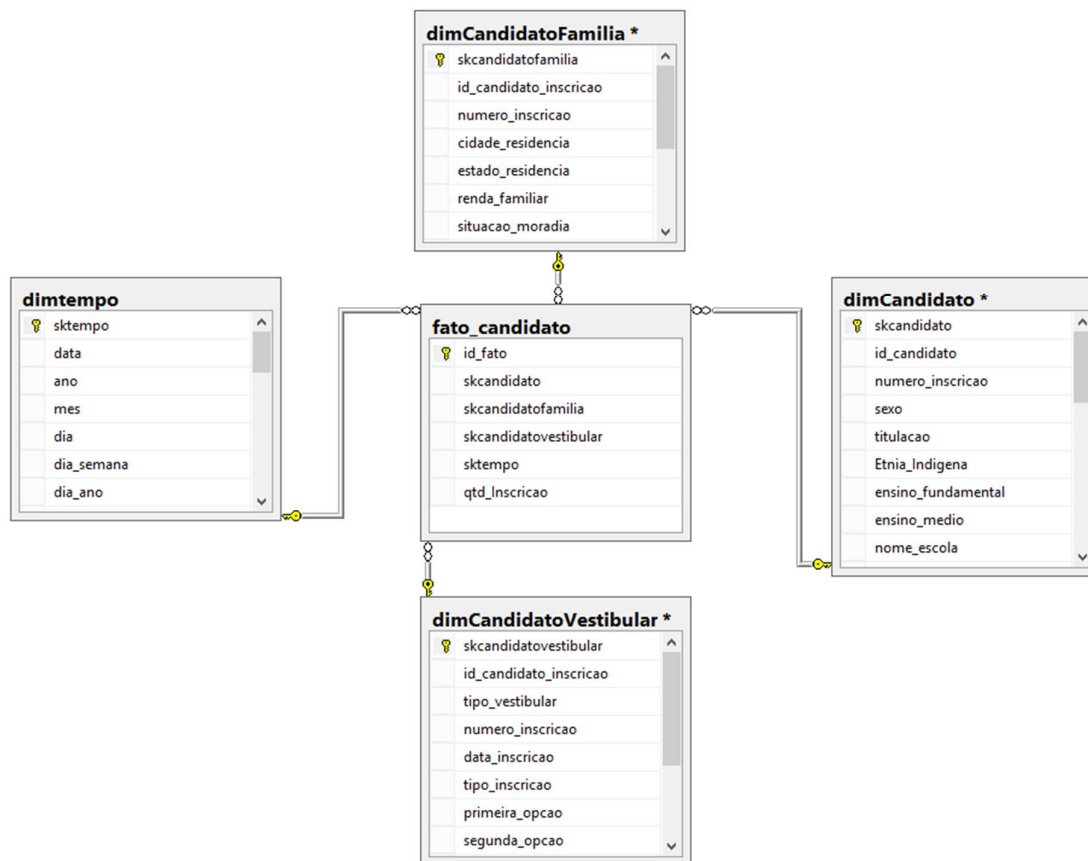
Dimensão Candidato_Família: Contém todas as informações referentes ao candidato relacionadas a família, tais como: renda familiar, se contribui com as despesas, cidade onde mora e etc.

Dimensão Candidato_Vestibular: Contém as informações referentes ao processo seletivo, tais como: cidade para realização da prova, primeira e segunda opção de curso, modalidade do curso e etc.

Como o processo seletivo é anual, a granularidade mínima é de uma inscrição válida por ano para cada candidato, de maneira que após a identificação da granularidade, a tabela fato foi definida com o totalizador da inscrição, é possível ter mais de uma inscrição para o mesmo candidato, no entanto cada uma delas representa apenas um processo seletivo.

Com todas as informações já definidas, foi utilizado o Microsoft SQL Server Management Studio para a criação física do banco de dados através do recurso de “*Diagramas de Banco de Dados*” foi realizado a modelagem do Data Mart, utilizando o esquema estrela, tendo a tabela fato central e correlacionada as dimensões conforme figura 6.

Figura 6 - Esquema “Estrela” – Modelagem das dimensões e tabela fato



11 PROJETO FÍSICO DO BANCO DE DADOS

Nessa etapa todas as estruturas físicas do banco são criadas, como a base foi criada na fase de modelagem, a criação das tabelas com seus campos e atributos foi feita utilizando os Diagramas de Banco de Dados, pois nele é possível criar e definir todas as restrições além das relações entre as tabelas e a definição dos atributos dos campos, não sendo necessário gerar scripts de implantação, pois a criação dos objetos é realizada naturalmente pelo Microsoft SQL Server Management Studio.

O SQL Server criar índices para todas as chaves primárias de todas as dimensões. Ao modelar foram definidas três dimensões com os dados dos candidatos divididos por núcleo, então existem as dimensões: Candidato, Candidato_Familia e Candidato_Vestibular, dessa maneira as pesquisas realizadas nos respectivos segmentos são processadas mais rapidamente, já que a quantidade de informação em cada uma delas é menor, apesar da mesma quantidade de registros.

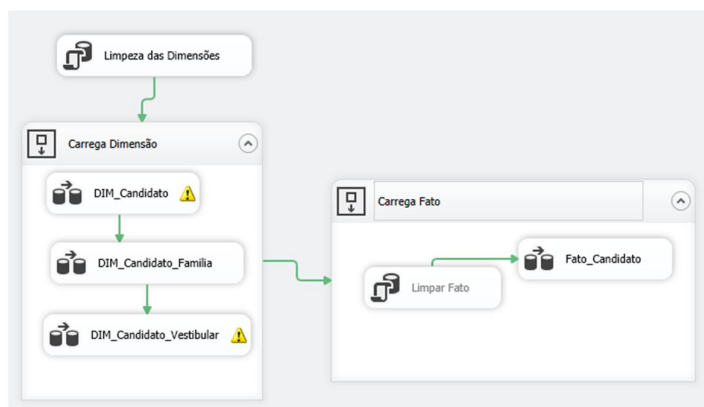
12 ETL – EXTRAÇÃO, TRANSFORMAÇÃO E CARGA

O processo de ETL é mais complicado, pois depende da origem das informações é nesse processo que acontece toda a seleção e integração dos dados, para essa fase foi utilizado o software da Microsoft *Integration Services* ele é o responsável pela parte de ETL da solução de BI escolhida.

O *Integration Services* possui diversos componentes gráficos que auxiliam na criação do processo de ETL, com ele é possível conectar as diversas fontes dos dados, realizar os devidos tratamentos e carregar para o Data Warehouse escolhido inserindo os dados já transformados nas respectivas dimensões, além disso como a tabela fato possui todas as chaves relacionais com as dimensões é a última a ser carregada, isso ocorre para garantir a integridade.

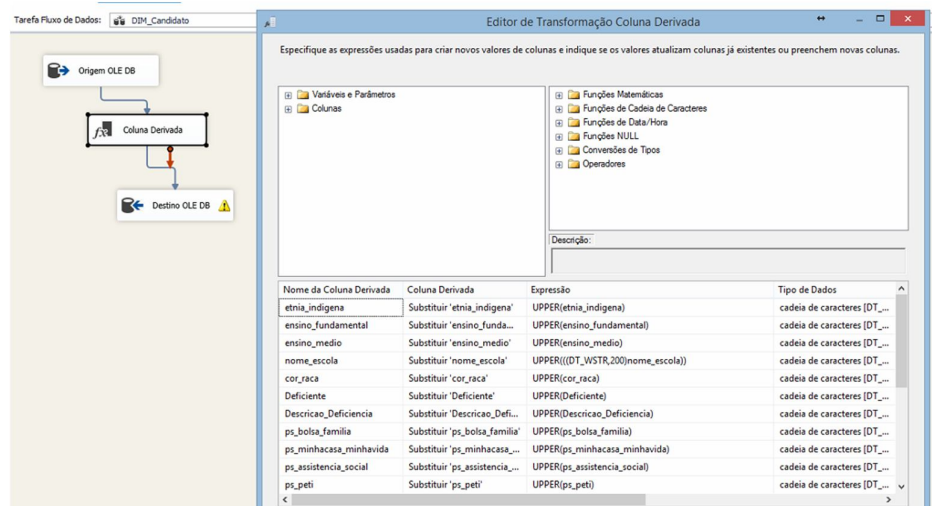
Como a origem dos dados é apenas um banco de dados transacional o processo de importação foi mais simples, pois foi criado na base de origem uma “View” com todos os dados necessários para o DM do Candidato e a partir disso foi realizado a extração dos dados, e a carga nas respectivas Dimensões. Após esse processo a tabela fato é então populada, conforme figura 7, o fluxo de importação e carga realizados no *Integration Services* são mostrados.

Figura 7 – Processo de Importação do *Integration Services*



Na figura 8 é exibido o processo realizado para carga e transformação da dimensão candidato.

Figura 8 – Processo interno do fluxo de importação da dimensão candidato

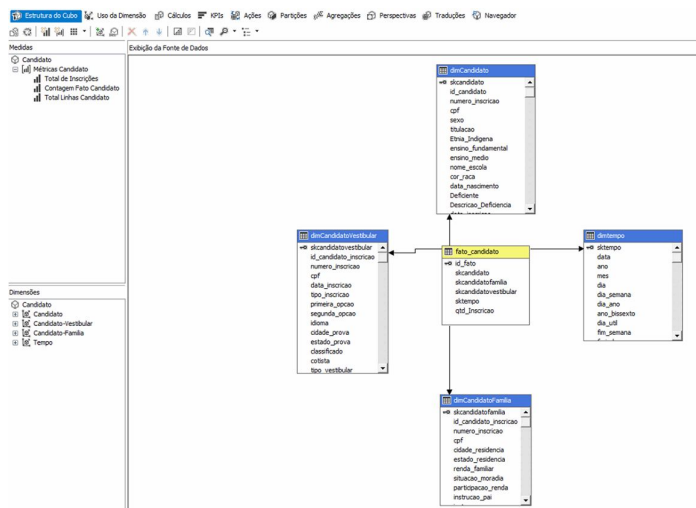


13 DESENVOLVIMENTO DE APLICAÇÕES

Após o processo de ETL realizado pelo *Integration Services* no passo anterior, o passo seguinte é a criação do cubo de dados através da ferramenta *Analysis Services* da Microsoft, com essa ferramenta é possível criar os cubos de dados a partir do Data Mart gerado nas etapas anteriores.

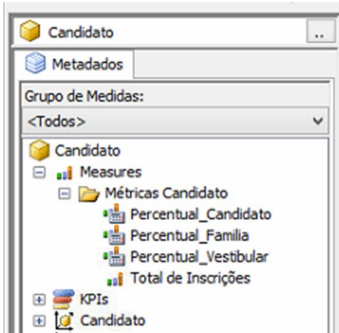
Foi criado um único cubo para a fato candidato conforme a figura 9:

Figura 9 – *Analysis Services* - Cubo fato_candidato



Já é possível através do ambiente do *Analysis Services* explorar o cubo após a sua criação e processamento conforme a figura 10 abaixo:

Figura 10 - Amostragem no último vestibular sobre gênero.



Dimensão	Hierarquia	Operador	Expressão do Filtro
Candidato	Cor Raca	Diferente	{ (Em branco), Unknown }
Tempo	Ano	Igual	{ 2014 }

Cor Raca	Percentual_Candidato
AMARELO(A) (DE ORIGEM ASIÁTICA)	0,00728560305513329
BRANCO(A)	0,179971080048941
INDÍGENA	0,00815690927292277
PARDO(A)	0,522190500908383
PRETO(A)	0,282062214971636

Apesar o ambiente do *Analysis Services* ser utilizado para a manipulação das dimensões e obtenção dos dados, por uma questão de usabilidade a ferramenta utilizada para a manipulação dos dados e posterior análise é o Excel que é bastante conhecida dos tomadores de decisão o que torna a curva de aprendizagem menor, permitindo uma maior velocidade na obtenção de resultados do projeto, além disso com a utilização do Excel não é necessário desenvolver nenhuma solução para a exibição dos dados.

Através do Excel, é possível realizar conexão com o *Analysis Services* e posteriormente, com a utilização de tabelas dinâmicas, manipular os dados escolhidos, (conforme figuras 11 e 12):

Figura 11 – Tabela dinâmica com amostragem no último vestibular sobre cor e raça

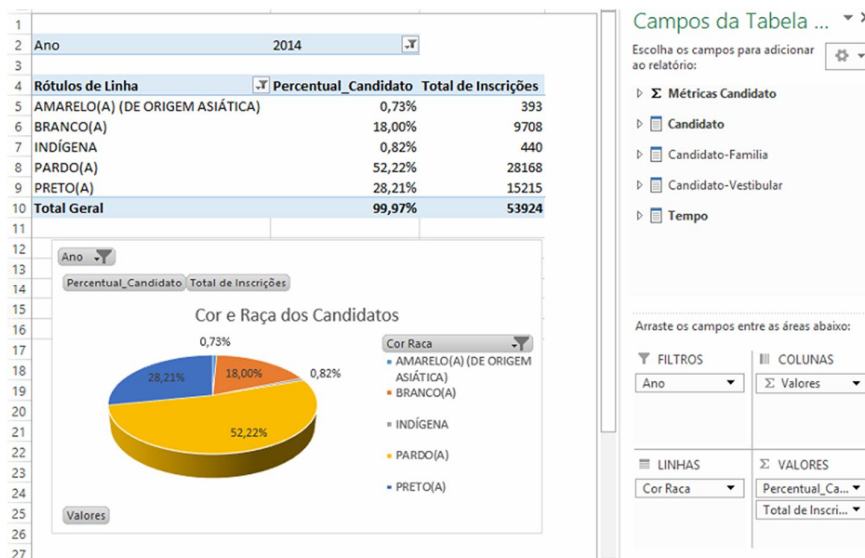
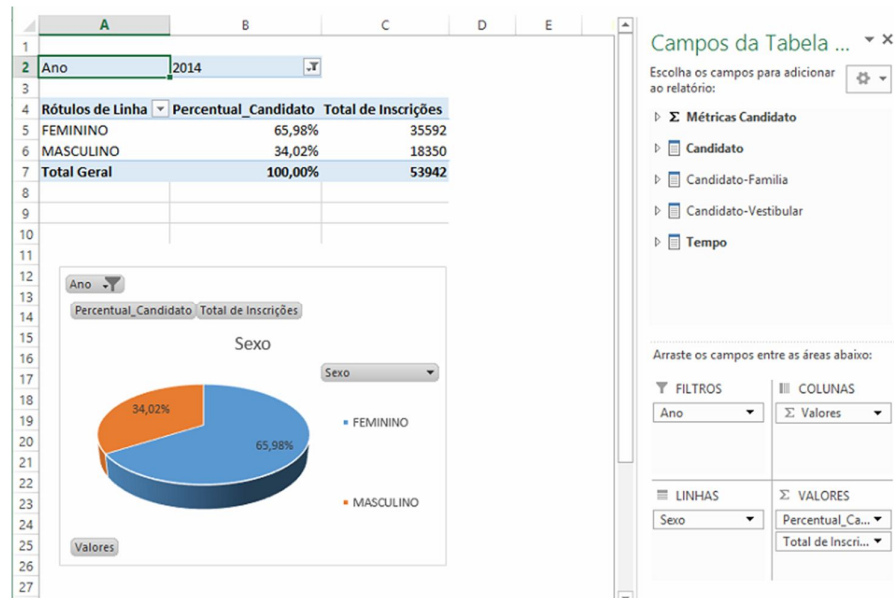
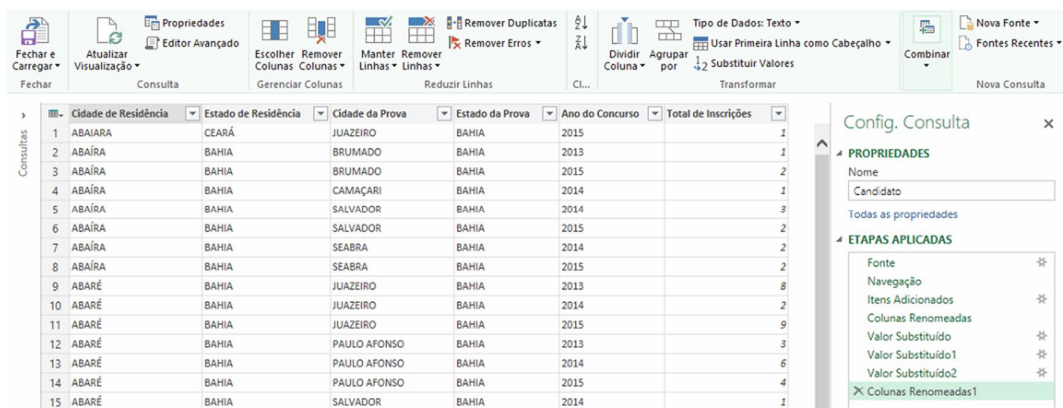


Figura 12 – Tabela dinâmica com amostragem no último vestibular sobre gênero



Além disso, o Excel possui um plugin chamado de *Power Map*, com esse plugin e com dados geográficos, é possível enriquecer mais ainda a apresentação dos dados, no entanto existe uma incompatibilidade com a estrutura multidimensional do Analyses Server, dessa maneira para que seja possível a utilização dos recursos do *Power Map*, é necessário utilizar o plugin Power Query também do Excel, dessa maneira o Power Query, é conectado a fonte de dados Multidimensional, realizando uma importação dos dados de maneira tabular para o Excel conforme a figura 13 abaixo:

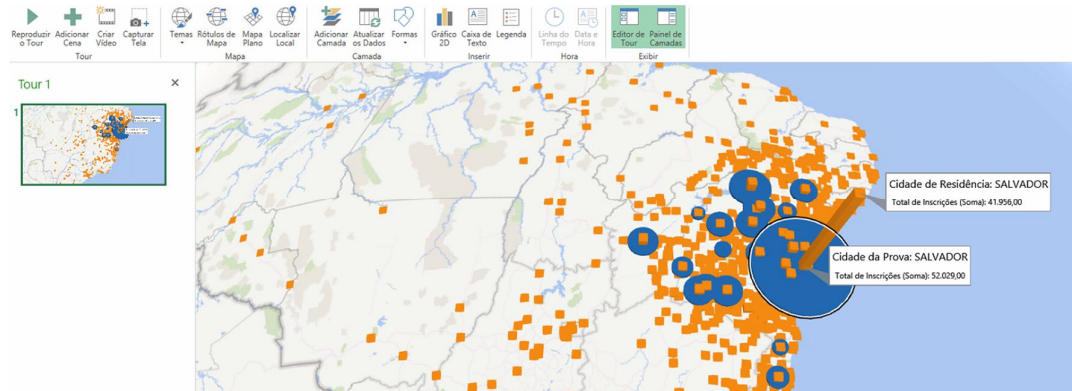
Figura 13 – Power Query



Em seguida os dados são transferidos para o *Power Map* onde é possível criar animações ricas para apresentar os mais diversos indicadores baseados em dados geográficos

(figura 14), onde é realizada uma apresentação comparando a quantidade de alunos que declaram a cidade de prova diferente da cidade de residência.

Figura 14 – Relação entre as cidades de prova e as cidades de residênciaz



14 HOMOLOGAÇÃO E IMPLANTAÇÃO

Apesar de ter sido realizado um recorte dos dois últimos anos com os dados dos candidatos do vestibular essa escolha foi acertada já que a informação está concentrada em uma única base de dados, dessa maneira o processo de modelagem, homologação e apresentação da ferramenta para os usuários foi bem direcionado pois a maioria das pessoas não conhecem a solução de BI. Dessa forma, a decisão de apresentar os recursos utilizando o Excel foi acertada, visto que é a ferramenta conhecida e muito utilizada para a geração de gráficos e agregações das informações sendo enriquecida e otimizada pela integração com o *Analysis Services*.

O processo de implantação será realizado posteriormente, pois o desenvolvimento deve seguir uma ordem de necessidade associada à disponibilidade das informações. O processo de homologação ocorre em paralelo, onde foi possível identificar diversas inconsistências nas bases transacionais que foram ajustadas pelas áreas responsáveis.

15 CONCLUSÃO

Mesmo com os avanços tecnológicos e a utilização de recursos cada vez mais especializados no processo de tomada de decisões, a integração das bases de dados e utilização de ferramentas que consolide esses dados criando informações estratégicas não é tão comum e simples. As organizações possuem diversas bases com informações espalhadas em ambientes distintos. Apesar disso existe uma busca pela qualidade da informação em

ambientes tão caóticos, as empresas começam a notar que essa integração é algo benéfico para a organização inteira visto que a informação pode ser compartilhada entre os departamentos alavancando os resultados.

Apesar de não ser um conceito novo, o BI ainda é pouco explorado pelas organizações, principalmente por conta do alto investimento para desenvolver a solução, pois além de infraestrutura e softwares é necessário o envolvimento de colaboradores externos para implementação da ferramenta e treinamento dos usuários internos.

Apesar da integração entre o *Analysis Services* e o Excel já existir a muito tempo, é notável que essa integração é pouco conhecida e utilizada, além disso o fato da Universidade estar passando por um processo de ajuste e adequação dos seus dados, tornou propício o desenvolvimento do ambiente, que vem sinalizando as inconsistências e problemas que estão sendo corrigidos, somado a isso, a disponibilidade das ferramentas da Microsoft que mostraram enorme potencial para o desempenho das funções com uma curva pequena de aprendizado.

Essa é a primeira etapa do projeto que vai continuar evoluindo inclusive como ferramenta de suporte acadêmica no desenvolvimento de ações voltadas para a comunidade científica interna. Após a implantação desta etapa, serão desenvolvidos os *data marts* relacionados aos campi, cursos, discentes e docentes de forma que a solução de BI possa atender a todas as áreas da Universidade, auxiliando no programa de reorganização e reorientação dos procedimentos acadêmicos.

REFERÊNCIAS

BALLARD, C; FARRELL, D; GUPTA, A; MAZUELA, C, VOHNIC, S. **Dimensional Modeling**: In a Business Intelligence Environment. IBM, 2006. 645 p.

GARTNER GROUP. Disponível em: <<http://www.gartner.com/it-glossary/business-intelligence-bi>>. Acesso em: 23 de mai. 2015;

IBM. Disponível em: <http://www.ibm.com/big-data/br/pt/big-data-and-analytics/?cm_mmc=always_on_-_bda_-_C212040W_-_definition_business_intelligence_business_intelligence&mkwid=sAQNbHqVI-dc_59894573424_432kjo9201>. Acesso em: 23 maio 2015.

KIMBALL, R, **Data Warehouse Toolkit**. São Paulo: Makron Books; 1998. 388 p.

KIMBALL, R, **The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling**, Third Edition. Indianápolis: Wiley; 2013. 388 p.

MACHADO F. N. R. **Tecnologia e Projeto de Data Warehouse**. 3. ed. São Paulo: Editora Érica; 2007. 318 p.

Nogare D, **Do banco de dados relacional à tomada de decisão**. São Paulo, Editora MVTech; 2014. 286 p. ISBN 9788565904094

REIS, E; TEIXEIRA, F; ARAÚJO, M. A. Implementando uma solução de Business Intelligence com o Microsoft SQL Server 2005 – Parte 1. **SQL Magazine**”, Rio de Janeiro, n. 59, p. 52-66, 2009.

SILBERSCHATZ, A; KORTH, H. F.; SUDARSHAN, S, **Sistema de Banco de Dados**. 5. ed. São Paulo, Editora Campus, 2006; 781 p.

THOMAS A. S. **Capital Intelectual**: A Nova Vantagem Competitiva das Empresas (em Português). 3. ed. São Paulo: Campus; 1998. 237 p.